

7ds

Set	Items	Description
S1	184213	DATABASE? OR DATABANK? OR DATAMIN? OR (DATA OR RECORD) (1W) - (MANAG? OR BASE? OR BANK? OR MINE? ? OR REPOSITOR? OR MINING) OR DB OR DBS OR OODB OR RDB OR DBMS OR RDBMS
S2	25878	(FAMILY OR FAMILIES) OR (GROUP? OR TABLE?) (3N) (RELATED OR - RELATIONSHIP? OR COMMON OR SIMILIAR OR KINDRED)
S3	28035	DUPLICAT? OR DUPE? ? OR DEDUPE? ? OR DEDUPLICAT?
S4	28	S1 AND S2 AND S3
S5	8	S4 AND IC=G06F?
S6	40490	(FAMILY OR FAMILIES) OR (GROUP? OR TABLE? OR SET OR SETS OR CLASS OR CLASSES) (5N) (RELATED OR RELATIONSHIP? OR COMMON OR - SIMILIAR OR KINDRED)
S7	36	S1 AND S6 AND S3
S8	15	S7 AND IC=G06F?
S9	7	S8 NOT S5

?show files

File 347:JAPIO Nov 1976-2004/May(Updated 040903)

(c) 2004 JPO & JAPIO

File 350:Derwent WPIX 1963-2004/UD,UM &UP=200460

(c) 2004 Thomson Derwent

?

?t s5/5/3,4,7

5/5/3 (Item 2 from file: 350)  
DIALOG(R)File 350:Derwent WPIX  
(c) 2004 Thomson Derwent. All rts. reserv.

015949937 \*\*Image available\*\*  
WPI Acc No: 2004-107778/200411  
XRPX Acc No: N04-085699

Duplicate data record identification method for personal computer,  
involves measuring co-occurrence of data in hierarchically related  
fields of tables

Patent Assignee: ANANTHAKRISHNA R (ANAN-I); CHAUDHURI S (CHAU-I); GANTI V  
(GANT-I)

Inventor: ANANTHAKRISHNA R; CHAUDHURI S; GANTI V

Number of Countries: 001 Number of Patents: 001

Patent Family:

Patent No	Kind	Date	Applicat No	Kind	Date	Week
US 20040003005	A1	20040101	US 2002186031	A	20020628	200411 B

Priority Applications (No Type Date): US 2002186031 A 20020628

Patent Details:

Patent No	Kind	Lan	Pg	Main IPC	Filing Notes
US 20040003005	A1		19	G06F-017/30	

Abstract (Basic): US 20040003005 A1

NOVELTY - The method involves providing multiple data records in one or more tables containing respective fields. The possible **duplicate** data records within the tables, are identified by measuring a co-occurrence of data in hierarchically **related** fields of the **table**.

DETAILED DESCRIPTION - INDEPENDENT CLAIMS are also included for the following:

(1) system for evaluating **duplicate** records in **database**; and  
(2) machine readable medium storing **duplicate** data record identification program.

USE - For detecting **duplicate** data records in **database** of data processing system such as personal computer and server connected to network such as internet, intranet, local area network and wide area network.

ADVANTAGE - The **duplicate** data records are accurately identified, using simple technique.

DESCRIPTION OF DRAWING(S) - The figure shows a flowchart explaining the **duplicate** data record identification method.

pp; 19 DwgNo 5/5

Title Terms: **DUPLICATE**; DATA; RECORD; IDENTIFY; METHOD; PERSON; COMPUTER; MEASURE; CO; OCCUR; DATA; HIERARCHY; RELATED; FIELD; TABLE

Derwent Class: T01

International Patent Class (Main): **G06F-017/30**

File Segment: EPI

5/5/4 (Item 3 from file: 350)  
DIALOG(R)File 350:Derwent WPIX  
(c) 2004 Thomson Derwent. All rts. reserv.

015684075 \*\*Image available\*\*  
WPI Acc No: 2003-746264/200370  
XRPX Acc No: N03-597946

Data family record managing method, involves adding designated record to family of records when determined that designated record is not duplicate of data records in family, and setting indicator to indicate relationship

Patent Assignee: MEINIG K (MEIN-I)

Inventor: MEINIG K

Number of Countries: 001 Number of Patents: 001

Patent Family:

Patent No Kind Date Applicat No Kind Date Week  
US 20030167253 A1 20030904 US 200291378 A 20020304 200370 B

Priority Applications (No Type Date): US 200291378 A 20020304

Patent Details:

Patent No Kind Lan Pg Main IPC Filing Notes  
US 20030167253 A1 39 G06F-007/00

Abstract (Basic): US 20030167253 A1

NOVELTY - The method involves adding a designated record to the potential **family** of records when it is automatically determined that the designated record is not a **duplicate** of the records in the **family**. An indicator is automatically set in each of the data records in the potential **family** of records to indicate a **family** relationship between the records.

DETAILED DESCRIPTION - INDEPENDENT CLAIMS are also included for the following:

(a) a computer-readable memory medium

(b) a **record management** system.

USE - Used for identifying relationship among data records.

ADVANTAGE - The method automatically determines missing information from a particular data record and also deletes **duplicate** records.

DESCRIPTION OF DRAWING(S) - The drawing shows an example flowchart of a routine for importing and automatically de-**deduplicating** a file of new data records.

pp; 39 DwgNo 7B/24

Title Terms: DATA; **FAMILY**; RECORD; MANAGE; METHOD; ADD; DESIGNATED;  
RECORD; **FAMILY**; RECORD; DETERMINE; DESIGNATED; RECORD; **DUPLICATE**;  
DATA; RECORD; **FAMILY**; SET; INDICATE; INDICATE; RELATED

Derwent Class: T01

International Patent Class (Main): **G06F-007/00**

File Segment: EPI

5/5/7 (Item 6 from file: 350)

DIALOG(R) File 350:Derwent WPIX

(c) 2004 Thomson Derwent. All rts. reserv.

013873764

WPI Acc No: 2001-357976/200138

XRAM Acc No: C01-111106

XPX Acc No: N01-260233

**Automated method for identifying related biomolecular sequences for features of interest from databases of nucleic acid and amino acid sequences**

Patent Assignee: ARS APPLIED RES SYSTEMS HOLDING NV (ISTF )

Inventor: COLINGE J; VAN HUIJSDUIJNEN R H; HOOFT VAN HUIJSDUIJNEN R

Number of Countries: 095 Number of Patents: 005

Patent Family:

Patent No	Kind	Date	Applicat No	Kind	Date	Week
EP 1103911	A1	20010530	EP 99811086	A	19991125	200138 B
WO 200138568	A2	20010531	WO 2000IB1676	A	20001116	200138
AU 200111697	A	20010604	AU 200111697	A	20001116	200153
EP 1232282	A2	20020821	EP 2000973154	A	20001116	200262
			WO 2000IB1676	A	20001116	
JP 2003515148	W	20030422	WO 2000IB1676	A	20001116	200336
			JP 2001539910	A	20001116	

Priority Applications (No Type Date): EP 99811086 A 19991125

Patent Details:

Patent No Kind Lan Pg Main IPC Filing Notes

EP 1103911 A1 E 16 G06F-019/00

Designated States (Regional): AL AT BE CH CY DE DK ES FI FR GB GR IE IT  
LI LT LU LV MC MK NL PT RO SE SI

WO 200138568 A2 E C12Q-001/68

Designated States (National): AE AG AL AM AT AU AZ BA BB BG BR BY BZ CA  
CH CN CR CU CZ DE DK DM DZ EE ES FI GB GD GE GH GM HR HU ID IL IN IS JP

KE KG KP KR KZ LC LK LS LT LU LV MA MD MG MK MN MW NX NO NZ PL PT  
RO RU SD SE SG SI SK SL TJ TM TR TT TZ UA UG US UZ VN YU ZA ZW  
Designated States (Regional): AT BE CH CY DE DK EA ES FI FR GB GH GM GR  
IE IT KE LS LU MC MW MZ NL OA PT SD SE SL SZ TR TZ UG ZW  
AU 200111697 A C12Q-001/68 Based on patent WO 200138568  
EP 1232282 A2 E C12Q-001/68 Based on patent WO 200138568  
Designated States (Regional): AL AT BE CH CY DE DK ES FI FR GB GR IE IT  
LI LT LU LV MC MK NL RO SI  
JP 2003515148 W 29 G01N-033/48 Based on patent WO 200138568

Abstract (Basic): EP 1103911 A1

NOVELTY - Automated method (A) for identifying related biomolecular sequences having defined features of interest from **databases**, is new. The **databases** comprise at least a first and a second set of sequences and each set is derived from a different type of organism.

DETAILED DESCRIPTION - Automated method (A) for identifying related biomolecular sequences having defined features of interest from **databases** comprises:

(1) establishing from the first set of sequences a non-redundant list of query sequences having the defined features of interest (first **family** members), using a **database** search program;

(2) performing sequence alignments with the first **family** members in a second set of sequences derived from a second type of organism, using a **database** search program and a preset similarity threshold, giving a list of second **family** members;

(3) establishing a two dimensional matrix displaying the first and second **family** members and their respective similarity values resulting from step (2), optionally displaying only those second **family** members having similarity values exceeding a preset threshold value;

(4) selecting from the matrix those pairs of first and second **family** members for which the similarity values are the best among all of the alignments that involve one of the two pair's members (orthologs).

USE - The method is used for identifying related biomolecular sequences having defined features of interest from **databases** of nucleic acid and amino acid sequences. The method can identify related or homologous sequences which can then be analyzed further in a laboratory environment. The features of interest can be specific classes of proteins or protein domains which can then serve as drug targets and are of high interest to researchers aimed at finding new drugs.

ADVANTAGE - The method is a high speed process for identifying the best-matching pair of orthologs. The basic parameters defining the sequences of interest and threshold values for the **database** searches can be set up to suit the needs of the researcher. Data can be analyzed off-line which saves time and costs.

pp; 16 DwgNo 0/2

Title Terms: AUTOMATIC; METHOD; IDENTIFY; RELATED; SEQUENCE; FEATURE; INTEREST; NUCLEIC; ACID; AMINO; ACID; SEQUENCE

Derwent Class: B04; D16; J04; T01

International Patent Class (Main): C12Q-001/68; G01N-033/48; **G06F-019/00**

International Patent Class (Additional): C12N-015/09; **G06F-017/30**

File Segment: CPI; EPI

?

9/5/6 (Item 5 from 350)  
DIALOG(R) File 350:Derwent WPIX  
(c) 2004 Thomson Derwent. All rts. reserv.

013595043 \*\*Image available\*\*

WPI Acc No: 2001-079250/200109

XRPX Acc No: N01-060293

**User accessible database organizing method for minimizing duplicate information amount, involves using data retrieval components for accessing index value and data entries**

Patent Assignee: BULL HN INFORMATION SYSTEMS INC (HONE )

Inventor: BEAUCHESNE R C

Number of Countries: 001 Number of Patents: 001

Patent Family:

Patent No	Kind	Date	Applicat No	Kind	Date	Week
US 6128626	A	20001003	US 98109118	A	19980630	200109 B

Priority Applications (No Type Date): US 98109118 A 19980630

Patent Details:

Patent No	Kind	Lan	Pg	Main IPC	Filing Notes
US 6128626	A		32	G06F-017/30	

Abstract (Basic): US 6128626 A

NOVELTY - A product directory index and data table structures for storing number of respective index and data entries are stored in tables (200-1-200-4) in **database** (200). A set of data retrieval components are provided in **database** for accessing index value and data entries and **database** is accessed via multiple client system to generate bill of material document for board assembly products.

DETAILED DESCRIPTION - The product directory index table structure storing index value entries coded in preset manner representing all board assembly products being currently manufactured is stored in a first table in **database**. Each index value entry contains key values having file key value coded for extracting data entries from other **tables** contained in **database** including information **related** to particular board assembly product and file version value coded for designating points within the **database** table where entries are added or deleted. Data table structures storing data entries containing different assembly board product related information used for generating bill of materials (BOM) documents used in manufacturing board assembly products is stored in other tables in the **database**. Each entry contains key values defining data selection criteria used in extracting entries utilized by particular board assembly product. The key values in each data entry includes ON and OFF field values which are set for identifying version of board products with and without data entry. The ON and OFF field values are used along with file version value for defining data selection criteria. The **database** is accessed via multiple client system having copy of data retrieval components to generate BOM document for board assembly products. An INDEPENDENT CLAIM is also included for software **database** system.

USE - For organizing **database** for minimizing storage amount of **duplicate** information used for generating bill of material documents for manufacturing printed circuit board assembly products.

ADVANTAGE - Enables operator or user to access **database** for obtaining information pertaining to particular PCB assembly.

DESCRIPTION OF DRAWING(S) - The figure shows components of **database** data retrieval software.

**Database** (200)

Tables (200-1-200-4)

pp; 32 DwgNo 2/6

Title Terms: USER; ACCESS; **DATABASE**; METHOD; **DUPLICATE**; INFORMATION;

AMOUNT; DATA; RETRIEVAL; COMPONENT; ACCESS; INDEX; VALUE; DATA; ENTER

Derwent Class: T01

International Patent Class (Main): G06F-017/30

File Segment: EPI

?